Hindawi
Journal of Electrical and Computer Engineering
Volume 2022, Article ID 9758584, 1 page
https://doi.org/10.1155/2022/9758584

Retraction

Retracted: Design of Customer Churn Early Warning System Based on Mobile Communication Technology Based on Data Mining

Journal of Electrical and Computer Engineering

Received 22 November 2022; Accepted 22 November 2022; Published 18 December 2022

Journal of Electrical and Computer Engineering has retracted the article titled "Design of Customer Churn Early Warning System Based on Mobile Communication Technology Based on Data Mining" [1] due to concerns that the peer review process has been compromised.

Following an investigation conducted by the Hindawi Research Integrity team [2], significant concerns were identified with the peer reviewers assigned to this article; the investigation has concluded that the peer review process was compromised. We therefore can no longer trust the peer review process, and the article is being retracted with the agreement of the editorial board.

References

[1] Q. Li, "Design of Customer Churn Early Warning System Based on Mobile Communication Technology Based on Data Mining," *Journal of Electrical and Computer Engineering*, vol. 2022, Article ID 9701349, 13 pages, 2022.

[2] L. Ferguson, "Advancing research integrity collaboratively and with vigour," 2022, https://www.hindawi.com/post/advancing-research-integrity-collaboratively-and-vigour/.

Hindawi
Journal of Electrical and Computer Engineering
Volume 2022, Article ID 9701349, 13 pages
https://doi.org/10.1155/2022/9701349

Research Article

Design of Customer Churn Early Warning System Based on Mobile Communication Technology Based on Data Mining

Qiang Li

School of Artificial Intelligence and Big Data, Zibo Vocational Institute, Zibo 255300, Shandong, China

Correspondence should be addressed to Qiang Li; 10657@zbvc.edu.cn

Received 9 April 2022; Revised 19 May 2022; Accepted 27 May 2022; Published 10 June 2022

Academic Editor: Xuefeng Shao

Customer churn is a fundamental problem faced by enterprises and an important factor affecting the operation of enterprises. Due to current market conditions and changing consumer behavior, it analyzes potential customer behavior trends by mining customer behavior data. This allows companies to set targets for looming market changes so that market movements can be predetermined. The rapid development of modern mobile communication technology makes the way of life need more new ways to adapt to the development of the new era. At the same time, with the rapid development of mobile communication technology, information management systems have been widely used. If a large amount of data can support decision-making information through data mining technology, it can drive the process of enterprise decision-making. It conducts purposeful and differentiated retention efforts on these customers. It increases the success rate of high-value customer retention, reduces the likelihood of customer churn, and reduces maintenance costs. It does this to achieve preset goals and minimize losses due to customer exit. This paper proposes and establishes a customer churn early warning system based on data mining. It uses this to find the customer trends behind a large amount of customer data. It uses the decision tree algorithm to participate in the decision-making process of the enterprise with this algorithm model. The RFT model proposed in the experiment and its results show that customer value is a key factor in the decision-making process of a firm. The accuracy rate is about 6% higher than that of the control group using the logistic regression model directly.

1. Introduction

With the rapid growth of China's economy and the intensification of competition with the commercial economy, domestic customer resources are faced with multiple choices. Even the inherent customer resources of some business units may be lost. This makes high-yield customers also become a competitive resource for many competitors.

In China, the serious loss of customers is a problem that enterprises have to face. For many companies, they are in some way grappling with the phenomenon of customer churn. In response to this phenomenon, it is necessary to thoroughly evaluate the internal and external factors that affect customer bias. With the wide application of data mining technology in life and work, the technology is introduced into the customer data analysis of enterprises, and the ability of computers to process big data can also be used. This thus enables the discovery of patterns behind customer data.

Oskouei R J believed that data mining had various techniques to extract valuable information or knowledge from data. At the same time, these techniques were applied to all data collected in all scientific fields [1]. In addition to the above data processing methods, Zhang *J* believed that with the increase in the amount of Internet of Things information, data storage management tends to be scattered, resulting in difficulties in data collaboration and interaction between sites, poor communication efficiency, and poor reliability. Therefore, it was considered that the application of blockchain technology was a major feature of supporting the modernization of management information technology [2]. Yang Q proposed a distance virtual data mining anomaly method based on FWSCA and differential evolution. In this way, the problems of inaccuracy, poor performance, and low efficiency of abnormal remote virtual communication data mining training can be solved [3]. The above was the application of data mining technology in various fields. Wang L N believed that spatiotemporal data mining based on network methods is beneficial for exploring the dynamic changes of mobile communication systems from a new perspective. A mobile communication system can be understood as a structure composed of interdependent base stations. The interaction between base stations can be evaluated by the similarity of base station data streams. The constructed network can reveal the interaction structure of human mobile communication activities [4]. Yumurtaci O proposed the theoretical framework of connectionism in the context of the relatively new networked social structure [5]. Kii A believed that Information and Communication Technology (ICT) was a powerful trigger for organizational change in all aspects and internal communication. He also studied the impact of ICT tool use on internal communication [6]. For the handling of customer churn, let *E*, B be proposed to determine the impact of words on customer churn based on the concept of usability heuristics. He also used decision tree diagrams and PLS modeling to determine which words had positive or negative effects [7]. Experimental results on real data collected by mobile show that the ensemble classifier based on key attributes has good performance in both classifier construction and customer churn prediction. The inadequacy of these studies above was that there was no combination of data mining and mobile communication technology to establish an effective early warning system. This cannot effectively utilize and integrate a large amount of customer data to avoid customer churn.

The innovations of this paper are that (1) For the mining of customer data, it presents the customer flow rules hidden behind a large number of customer data in a relatively complete system form. (2) The use of mobile communication technology based on data mining technology enables the flow of customer data to be better controlled to analyze the tendency of customer flow. This gives enterprise companies a head start in customer management. (3) It narrates the role of mobile communication technology in data generation, making the data source of customer loss more clearly displayed.

2. Application of Data Mining Mobile Communication Technology in Customer Churn Early Warning

2.1. Data Mining Methods. Data mining is the process of extracting information and knowledge hidden in data. Data mining methods are usually divided into two types: direct data mining and indirect data mining. It applies data mining technology to complete various work contents, including classification, evaluation, and prediction. However, different data mining methods also have different algorithms [8].

2.1.1. The Basic Steps and Operation Process of Data Mining Technology. American scientists first proposed the data interactive mining process model [9]. Its basic principle is shown in Figure 1:

It can be seen from Figure 1 that the interactive process model of data mining is usually divided into application scenarios of data mining technology. It applies data mining technology to transform data sources. It then aggregates the obtained useful information to generate the decision to be generated and finally evaluates the effectiveness of the model. The steps of data mining include defining the problem, establishing a data mining library, analyzing the data, preparing the data, establishing the model, evaluating the model, and implementing it.

2.1.2. Establishment of Data Mining Model for Customer Churn. Businesses are run in pursuit of profit. Therefore, in a broad sense, as long as the customers that cause the loss of corporate profits can be regarded as lost customers [10]. Through the integration of a large amount of customer data, it needs to choose a specific data mining model to play a key role in the operation of the enterprise. Here, we choose to use the decision tree algorithm in operations research for modeling. This is used to judge the credit degree of the lost customers, which will play a role in the decision-making of enterprises [11]. The following is the basic process of CAMM (Classification Algorithm by minsup and minconf) decision tree algorithm. The concept of information gain is an indicator used to measure the difference between two probability distributions P and Q. The requirement of the decision tree algorithm is to select the decision attribute as the current node to provide the maximum information gain to minimize the branches of the decision tree. We calculate the information gain of decision attribute by the basic attributes of decision (predictability, selectivity, and subjectivity).

The set of input samples is represented by Q, and the set of Q samples has n independent values. It uses $V_i, i = 1, 2, ...n$, where n represents n classes. This assumes that T_i is a subset of V_i and is an element in Q, using t_i to denote the set of element numbers of T_i. The amount of information gain of the set Q can be expressed as the following formula:

$$I\left(t_1, t_2, \ldots, t_n\right) = -\sum_{i=1}^{n} p_i \log_2\left(p_i\right). \tag{1}$$

In formula (1), $p_i = t_i / |Q|$, and $|Q|$ represents the number of tuples in the training sample dataset. Here, it is assumed that the value of attribute B is $\{b_1, b_2, \ldots, b_m\}$, and there are m different values. According to the different values of m, partition Q can be divided into m subsets. We use q_j to denote the subset corresponding to element b_j in attribute B, where $j = 1, 2, \ldots, m$. This assumes that B is selected as the decision attribute, and the branches of the decision tree correspond one-to-one with these subsets. This assumes that the number of elements belonging to V_i in subset q_j is represented by q_{ij}, then attribute B corresponds to the expected amount of information of class V_i. That is, the entropy corresponding to the attribute can be expressed as follows:

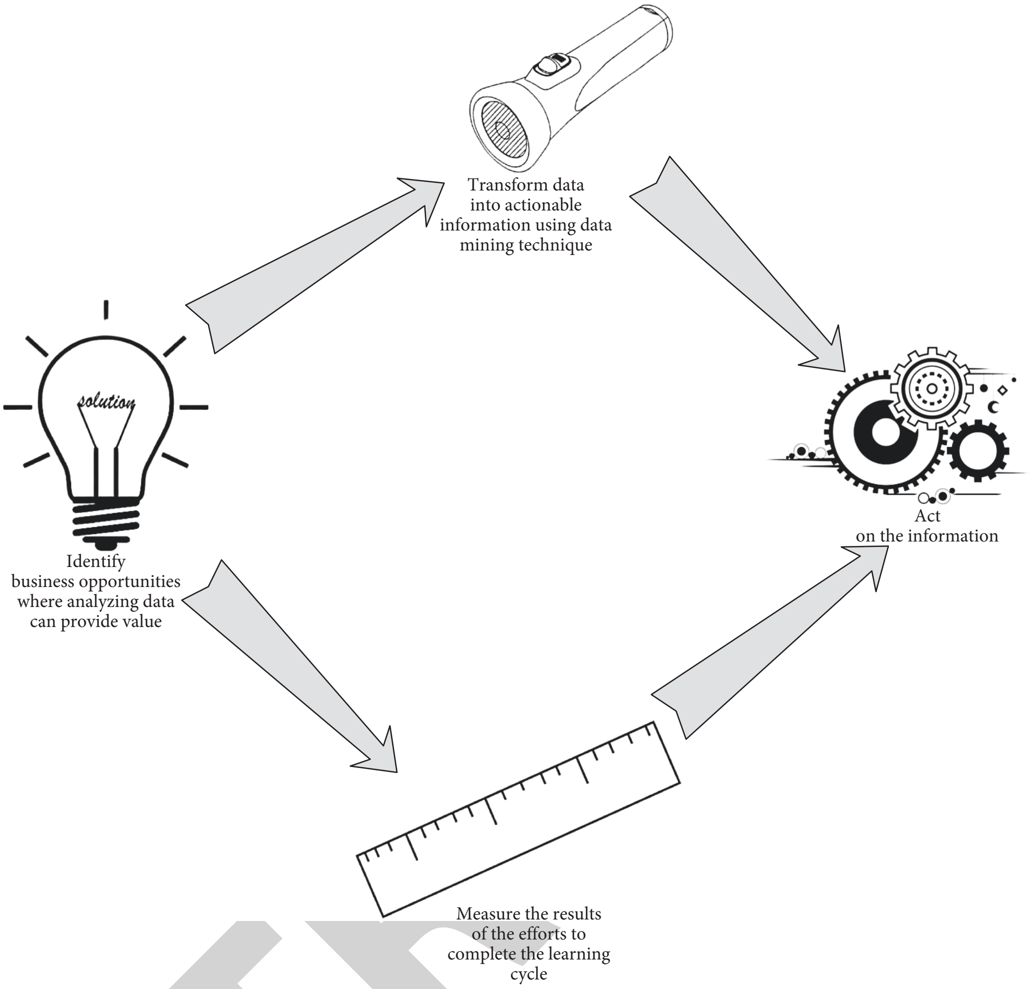

Figure 1: Data mining interactive process model.

$$F(B) = \sum_{j=1}^{m} \frac{q_{1j} + q_{2j} + \dots + q_{nj}}{|Q|} I\left(q_{1j}, q_{2j}, \dots, q_{nj}\right). \tag{2}$$

Here, w_j is used to represent the weight of subset q_j in the data Q, that is, $w_j = q_{1j} + q_{2j} + \dots + q_{nj}/|Q|$. The entropy $I\left(q_{1j}, q_{2j}, \dots, q_{nj}\right)$ of each element of attribute B for class V_j can be given by the following formula, namely:

$$I\left(q_{1j}, q_{2j}, \dots, q_{nj}\right) = -\sum_{i=1}^{n} p_{ij} \log\left(p_{ij}\right). \tag{3}$$

In formula (3), p_{ij} represents the weight of subset q_j for classes V_i and $i = 1, 2, \dots, n$, that is, $p_{ij} = q_{ij}/|q_j|$.

According to formulas (1)–(3), the expression of information gain can be obtained when the selection attribute B is the decision attribute.

$$\text{Gain}(B) = I\left(t_1, t_2, \dots, t_n\right) - E(B). \tag{4}$$

The above calculation formula calculates the information gain of attribute B and finally selects the attribute with the largest information gain as the decision tree node of Q [12]. It is assumed that the minimum support of each decision attribute node is specified by a threshold of β, and its value range is $[0, 1]$. This assumes that the attribute nodes of the decision tree are as follows: decision attribute 1 is A, decision attribute B is 2,..., decision attribute k is P, $1 \leq k < m$, and m represents the number of decision attributes.

Assuming that Y is the set of elements corresponding to the above attribute nodes, X_i is used to represent one of the subsets, and the category identification attribute of X_i is V_i, where $1 \leq i \leq n$ is used. If the number of tuples in X_i accounts for q% of the total number of Y tuples in the set, then q% is called the support degree of X_i for Y.

The above is the establishment of decision tree rules. It then assumes that the minimum confidence threshold is $(0 \leq \mu \leq 1)$, and its value range is the same as β. It is assumed that the limit that the rule can be adopted is consistent with the node of the decision tree attribute, where m is the condition of a classification rule generated by the decision tree. If the sample elements whose attribute values $V_i \, (1 \leq i < n)$ are in the set Y occupy q%, q% is the reliability V_i of the pair in the set Y. In the final result of the decision tree algorithm, the branches with a minimum confidence of less than 90% and a minimum support of less than 10% are generally not used.

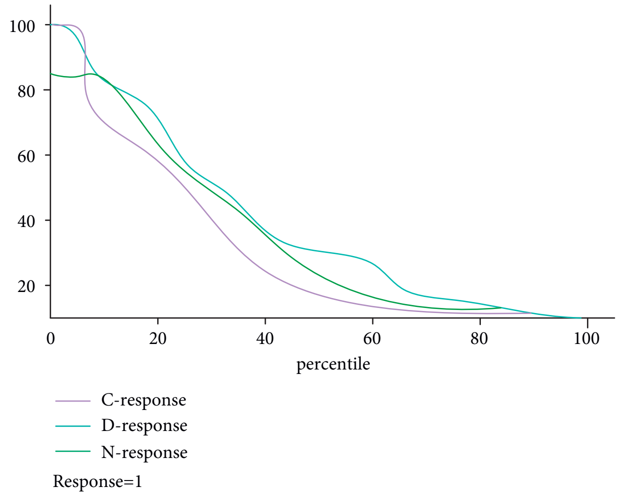

Figure 2: Evaluation diagram.

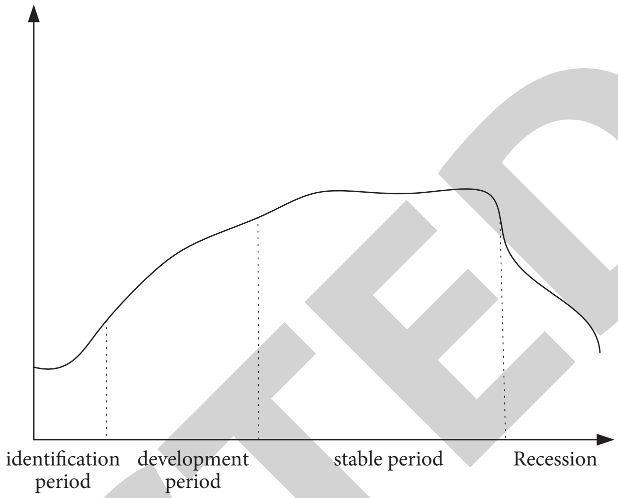

Figure 3: Typical customer lifecycle.

2.1.3. Evaluation of Data Mining Models. When data mining progresses to the last part, it is necessary to evaluate the validity of the model through inspection tools such as profit and loss statement tools and error judgment tools [13]. When evaluating the test, the choice of data will also have a certain impact on the results. Here, the data that has not participated in the decision tree algorithm is used for testing. The test data here will use a mixed test of the data set, with and without feedback. Decision tree algorithms include Algorithms 5.0, C4.5, and CART. In addition, to highlight the advantages of the decision tree algorithm (That is, a method to describe the decision problem in a table, and this table is also called a decision matrix.) used, it introduces a decision table (represented by D) and a neural network (represented by N) for comparison and analysis with the decision tree model. The evaluation chart comparison is shown in Figure 2:

The results of the above three models are analyzed as follows. Among them, the correct rate of the decision tree is 89.68%, while the correct rate of the decision table is 11.52%, and the correct rate of the neural network is 81.02%. To sum up, the effect of decision tree C5.0 is better [14].

2.2. Concepts Related to Customer Churn and How to Deal with Them

2.2.1. Customer Life Cycle Management. For businesses, customer churn and business-to-customer lifecycle management are closely related. A typical customer life curve is shown in Figure 3.

The horizontal axis of Figure 3 is the time variable. The vertical axis is a two-dimensional curve model diagram established as a parameter representing the customer value level. For different enterprises, the trend of the customer life cycle is not the same. There are also transitions and jumps at different stages of the customer relationship, which do not necessarily require four processes, as shown in Figure 3 [15]. However, the customer relationship status will show a certain regularity in the characteristics of the life cycle, so the life cycle management has certain controllability for the management of the customer relationship.

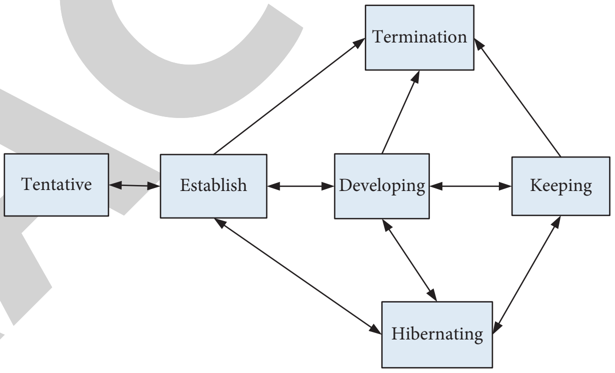

Figure 4: State model of customer relationship development process.

There is also a certain process relationship in the transformation and development of customer relationships, as shown in Figure 4.

The customer life cycle is a critical part of an enterprise's customer management process. Its effective analysis and management of the customer life cycle play an important role in preventing customer churn.

2.2.2. Establishment of Customer Churn Concept and Early Warning Model. For the definition of customer churn, there are generally two cases [16]. The first is customers who churn actively, and the second is customers who churn passively. Both are due to some factors. They choose to buy products from other companies in the process of cooperating with companies. It is just that the former is the self's active choice, and the latter is the passive release of the cooperation by the enterprise during the cooperation process. Therefore, we will design the structure of the churn early warning model according to the characteristics of customer churn-related data. The new model structure shown in Figure 5 is expected to obtain better results in customer churn prediction.

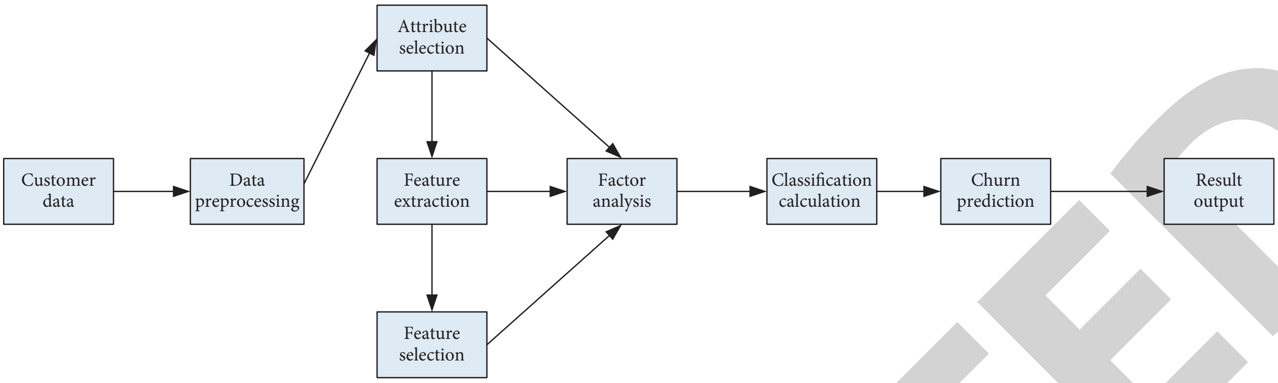

Figure 5: Model diagram for customer churn prediction.

2.2.3. Construction of Customer Churn Early Warning System Structure. The whole process includes customer data preprocessing churn rate calculation churn prediction result output. It first needs to establish a data warehouse for customer data. Then according to the existing data, it performs attribute selection and reduction through data association analysis and data mining methods. It extracts customer feature vectors and builds a predictive model expert system [17]. Its main flow chart is shown in Figure 6.

Next, it establishes a customer churn prediction model for two types of errors, as shown in Figure 7. The first type of error is generally an error with a large loss. That is, the error of the "churn" type is judged as the "nonchurn" type as the first type of error. The second type of error is to misjudge the "nonchurn" type as the "churn" type [18].

Loss function calculation is a function that maps the value of a random event or its related random variables to nonnegative real numbers to represent the "risk" or "loss" of the random event. Complexity is the complexity of a thing and can be measured by the length of the computer language required to describe it. The following is a brief description of loss function calculation, complexity calculation, attribute dimension calculation, satisfaction evaluation, and model interpretation. The following are the relevant expressions for the two types of errors, and the loss function is defined as follows:

$$\begin{aligned} l_n &= \frac{(W_{\max} - W)}{(W_{\max} - W_{\min})}, \\ W &= H_1 * Q_1 * T_1 + H_2 * Q_2 * T_2. \end{aligned} \tag{5}$$

In the above formula, H_1 is the proportion of positive classes in the training set, H_2 is the proportion of negative classes in the training set, and Q_1 is the loss caused by misclassifying a positive class into a "negative" class. Q_2 is the loss caused by misclassifying a "negative" class into a positive class, 1 is an error rate of T_1, and 2 is an error rate of T_2. Its definition complexity function is as follows:

$$l_b = g(b) = \frac{(B_{\max} - B)}{(B_{\max} - B_{\min})}, l_b \in [0, 1]. \tag{6}$$

If $B = B_{\max}$, then $l_b = 1$, $g(b)$ is the complexity function, $B_{\min}$ is the minimum complexity, and 3 is the maximum complexity.

This sets the dimension of the attribute set to be s and $s_{\max} \neq s_{\min}$, then the attribute dimension function is defined as follows:

$$l_s = f(s) = \frac{(s_{\max} - s)}{(s_{\max} - s_{\min})}, l_s \in [0, 1]. \tag{7}$$

Among them, $f(s)$ is the attribute dimension function, $s_{\max}$ is the dimension of the original attribute set, and $s_{\min}$ is the minimum dimension of the extracted attribute set. If $s_{\max} = s_{\min}$, then $l_d = 1$. The introduction of attribute dimension calculation solves the problem that most model evaluations need to determine the dimension of the selected attribute set in advance through experience and need to try different dimensions many times. It also provides a reference for interpreting predictive models [19]. This defines the satisfaction evaluation function as $\Phi(l_n, l_b, l_s)$, which is only related to 3 factors l_n, l_b and l_s. Here, a linear weighting function is taken, that is,

$$\Phi(l_n, l_b, l_s) = \frac{(\mu_n l_n + \mu_b l_b + \mu_s l_s)}{(\mu_n + \mu_b + \mu_s)}, \tag{8}$$

where μ_n, μ_b, μ_s is the weighting factor of l_n, l_b, l_s, respectively. In practical research, model evaluation is often a satisfactory optimization problem in essence, and the obtained solution is also a satisfactory solution.

2.3. Application of Mobile Communication Technology in Customer Life Cycle Management

2.3.1. Mobile Communication Technology. With the growing economy and the continuous improvement of informatization in people's daily life, the penetration of informatization in lives makes the way of life need more new ways. Therefore, modern mobile communication technology has developed rapidly, from 1G to 4G, and the ongoing 5G. As shown in Figure 8, the following will introduce the GPRS network architecture [20]. By adding SGSN and GGSN, the former is a GPRS service support node and the latter is a gateway GPRS support node, which is of great benefit to the transmission of high-speed data.

The rapid development of mobile communication technology has accelerated the flow of information. And this flow of large amounts of data information is the potential

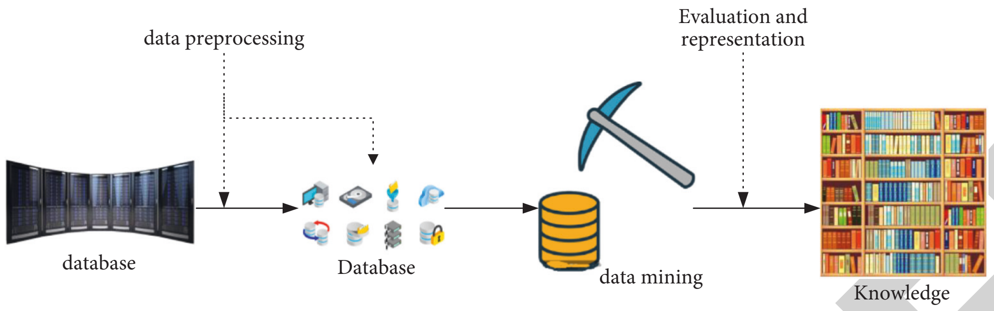

Figure 6: Data mining flowchart.

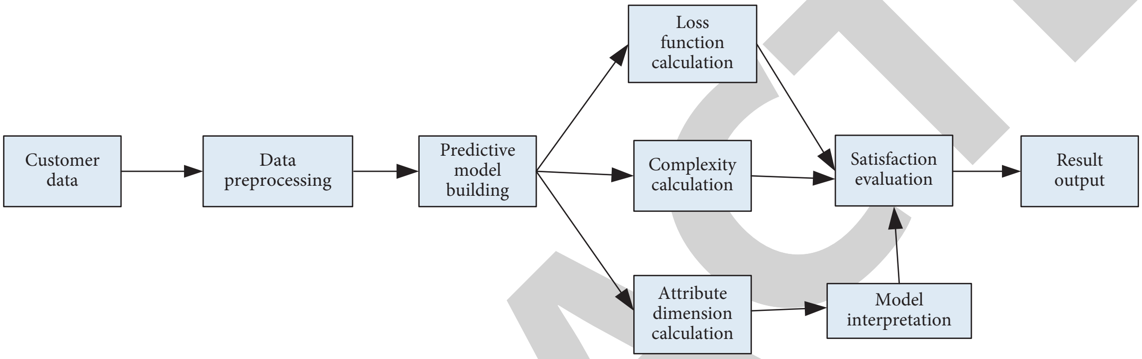

Figure 7: Two types of wrong customer churn prediction model.

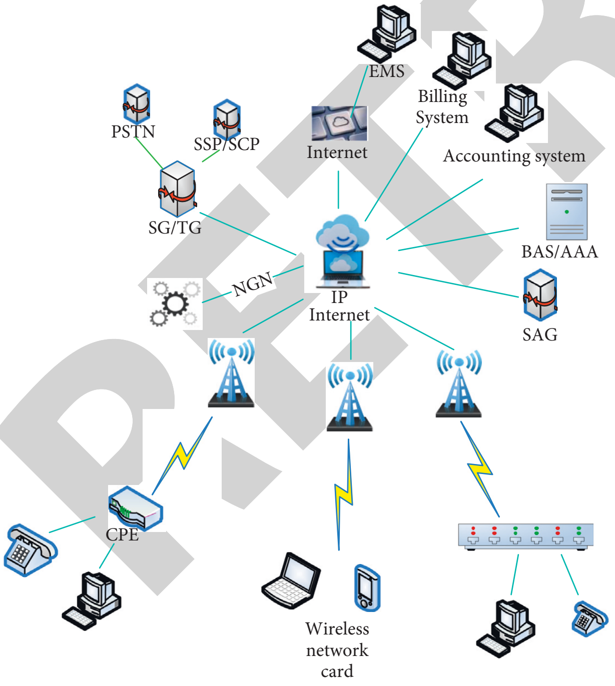

Figure 8: GPRS network structure diagram.

information of each enterprise-related customer. Mobile communication technology has a large role in promoting the generation of customer data, which provides a data source for the establishment of an early warning system for customer churn.

2.3.2. Customer Data Processing Based on Mobile Communication Technology. The RFT model referenced below is a model applicable to the telecommunications industry. This model adds some new definitions to the RFM model. Because the three factors R (Recency), F (Frequency), and M (Monetary) involved in the RFM model cannot well analyze and explain customer behavior, and because the value range of the commodity base price is very large, it will fail to predict.

Definition 1. Total profit T: The basic operating cost of the industry is A, the basic price of the package is Q, and the user payment time matrix is F.

$$T = \frac{F * (Q - A)}{A}. \tag{9}$$

Definition 2. The efficiency model of user purchase is RFT: R, F, and T are the three indicators of the above model. The matrix acquisition method corresponding to the three indicators R, F, and T is directly from the database or through the main conversion method.

A represents the basic operating cost matrix of the telecom industry, the basic service fee that the telecommunications company needs to provide to the user every day this month. A_{ii} represents the operating cost paid by the user to the user on the ith day in matrix A, where A_{ii} is the stage we can use to calculate. In the general case $A_{11} = A_{22} = \ldots = A_{mm}$, that is, we assume that the cost is constant. Matrix A is shown in the formula:

$$A(m,m) = \begin{pmatrix} A_{11} & 0 & \cdots & 0 \\ 0 & A_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_{mm} \end{pmatrix}. \tag{10}$$

This set the matrix formed by the user to pay extra fees in addition to the package this month as E, E_{ii} means that the user in matrix E needs to pay extra fees on the ith day of the month and E_{ii} can be directly obtained from the data system. If the package fee is not exceeded, the value is 1, and the user's additional payment matrix E can be represented by the formula:

$$E(n,n) = \begin{pmatrix} E_{11} & 0 & \cdots & 0 \\ 0 & E_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & E_{mm} \end{pmatrix}. \tag{11}$$

This sets the matrix formed by the basic package cost of the user this month as P. Since Telecom's current package is in the form of daily deductions to charge customers' monthly rental fees, all $P(i,i)$ are the daily rental fees deducted by the user on the ith day. The matrix P can also be obtained directly through the data system, and the user basic package matrix P can be expressed by the formula:

$$P(n,n) = \begin{pmatrix} P_{11} & 0 & \cdots & 0 \\ 0 & P_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & P_{mm} \end{pmatrix}. \tag{12}$$

We can easily get R from database tables. When using R, F, and T to calculate, if it wants to get accurate results, it needs to normalize the data. This can be done by using a normalization transformation to process the data matrix [21].

Correlation analysis models measure the linear correlation between variables by calculating the Spearman correlation coefficient between the variables.

$$r = \frac{\sum\left(R_{xi} - \overline{R}_x\right)\left(R_{yi} - \overline{R}_y\right)}{\sqrt{\sum\left(R_{xi} - \overline{R}_x\right)^2 \cdot \sum\left(R_{yi} - \overline{R}_y\right)^2}}. \tag{13}$$

Among them, R_{xi} and R_{yi} represent the sorted order of the ith, x variable, and y variable, respectively, and $\overline{R}_x$ and $\overline{R}_y$ represent the mean value R_{xi} and R_{yi} of the sum.

As a rule of thumb, different |r| values indicate different degrees of linear correlation:

|r|<0.1 means weak correlation; 0.1 ≤|r|<0.3 means low linear correlation; 0.3 ≤|r|<0.5 means low-to-medium linear correlation; 0.5 ≤|r|<0.8 means moderate linear correlation;

0.8 ≤|r|<1.0 means a highly linear correlation.

2.3.3. Application of Data Mining Mobile Communication Technology in Customer Data. The purpose of call data preprocessing is to generate corresponding contact sequences from call log files.

It can be represented by the following formula:

$$N^{\mu} = \sum_{t=t_{\mu}}^{t_{\mu+1}-1} D_{ij}^{t}, t \in \left[t_{\mu}, t_{\mu+1} - 1\right]. \tag{14}$$

The matrix N^{μ} removes self-loops and multigraphs in the process of compressing the time dimension. Password settings for mobile applications are one of the important components of mobile security. The effectiveness of the password against attacks can be objectively expressed by the password strength. Password strength can be expressed as a function of password length, complexity, and unpredictability. Password strength can be represented by the following function f:

$$f\colon X^{*} \longrightarrow \mathrm{Q}. \tag{15}$$

Among them, function f represents the correspondence between the input/output of a certain password cracking system. $X*$ a means any set of strings on a character table X. The output value Q is a certain score s calculated for the corresponding input.

The most common way to increase the strength of a password is to increase the password length. Studies have shown that a large number of users use single characters for their passwords. In addition to the password length, it is also related to the complexity and unpredictability of the password. What is taken here is the effect of increasing the password length on the password strength. This greatly reduces the security of the system. The relationship between the password strength value t and the password length m can be expressed as follows:

$$t = c^{m}. \tag{16}$$

Among them, m represents the character set size of the password candidate. It can be seen from formula (15) that increasing the password length can effectively enhance the password strength.

Increasing the complexity of passwords is here to increase character type analysis, which is another common method of strengthening passwords. Customers can use numbers (D), lowercase letters (P), uppercase letters (Z), and special characters (W) for password combinations. The following is the number of password types:

When using only numbers:

$$D_n = 10^{n}. \tag{17}$$

When using only lowercase letters:

$$P_n = 26^{n}. \tag{18}$$

When only letters are used, uppercase and lowercase letters are included:

$$C_n^{y} P_y \cdot C_{n-y}^{n-y} Z_{n-y} = C_n^{y} 26^{y} \times 26^{n-y} = C_n^{y} 26^{n}. \tag{19}$$

When containing numbers, lowercase letters, uppercase letters, and special symbols:

Table 1: Number of set tuples based on user attributes.

Property set	Attribute value	Number of tuples	Number of tuples	Number of tuples	Number of tuples
Stat ID attributes	Customer status in January 2009	Stat = 0 9956	Stat = 10. 421	Stat = 5 204	Stat = space 870
Create_Time.	Network access time ≥3 years	5888	102	50	153
	1years ≤ network access time <3 years	1520	122	42	140
Credit_Level	Credit rating = advanced {6,7,8,9, 10}	1156	52	26	22
	Credit rating = advanced {1,2,3,4, 5}	8728	411	175	841
ZD-5/previous	Consumption in December ≤60yuan	355	45	30	650
Consumption	Consumption in December >60yuan	9500	382	180	240
ZD-10 ZD-11	Average consumption in the first three months ≤60yuan	140	20	10	240
ZD-12	Average consumption in the first three months >60yuan	9800	395	188	476

Table 2: Customer churn profile.

	Total application	Unusual use	0 call charges	Under the call 40% drop	Lost	Churn rate (%)
Training set	32000	1030	7	2250	3300	10.3
Test set	31300	960	12	1669	2607	8.5

Note. "Normal use" includes user suspension, user disassembly, and double-stop for arrears.

$$C_n^x D_x \cdot C_{n-x}^y P_y \cdot C_{n-x-y}^z Z_Z \cdot C_{n-x-y-z}^{n-x-y-z} W_{n-x-y-z} = C_n^x C_{n-x}^y C_{n-x-y}^z 10^x \cdot 26^{y+z} \cdot 32^{n-x-y-z}. \quad (20)$$

3. Experiment Design and Results Analysis

3.1. The Process of Training the Model. The training data set is $H = 16335$ user records in a certain area, and the H attribute selects the state customer status as the attribute class identifier, integrates it into a standard data format according to the data input rules, and uses the search function of the database to count the training data as shown in Table 1.

In Table 1, users are divided into online state $B_1 = 0$, single-stop state $B_2 = 10$, full-stop state $B_3 = 5$, and empty space churn state $B_4 = 15$ in the current month. Among them, the number of class B_1 tuples is $t_1 = 9956$, the number of class tuples of B_2 is $t_2 = 421$, the number of class tuples of C_3 is $t_3 = 204$, and the number of class tuples of C_4 is $t_4 = 870$.

3.2. Customer Attribute Results and Characteristic Analysis of Customer Churn. Table 1 is a record of the number of calls made by customers in a month and the cost of calls per month. It finds that the consumption attribute gain value of the previous month is the largest by comparing the gain value of each attribute. Therefore, the decision point for selecting the next layer is the consumption situation of the previous month. Table 2 is an overview of customer churn.

It can be seen that the overall churn rate is around 10%. Among them, customers with abnormal use account for about 1/3, customers with a 40% decrease in call charges account for about 2/3, and customers with zero call charges are relatively few.

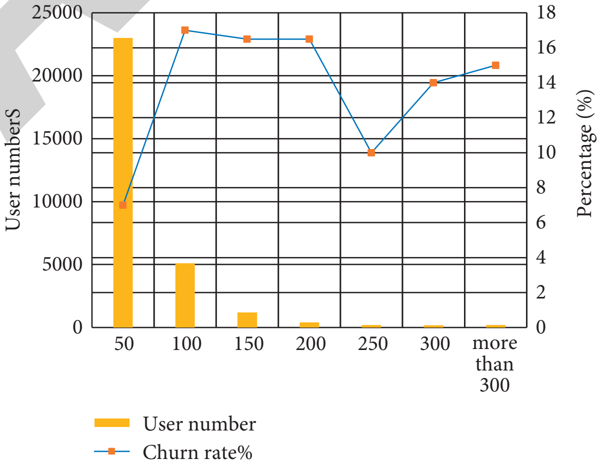

Figure 9: Churn vs Average Revenue Per User.

As shown in Figure 9, users with an average revenue per user between 0 and 50 account for a considerable number of users. However, the churn rate of this part of users is 7.7%, which is lower than the average churn rate of 10.3%. On the contrary, users with an average revenue per user between 50 and 200 showed a churn rate higher than 16%, which is more worthy of attention.

In addition to the above-related factors, the loss of customers also has a certain relationship with the proportion of intraregional call costs, the proportion of long-distance call costs, the proportion of information cost, and the number of IP calls. Figure 10 shows the relationship of the four factors to the churn rate.

As can be seen in Figure 10, the number of users with different monthly rental fees is generally distributed, normally around 40%. The churn rate gradually decreases as the

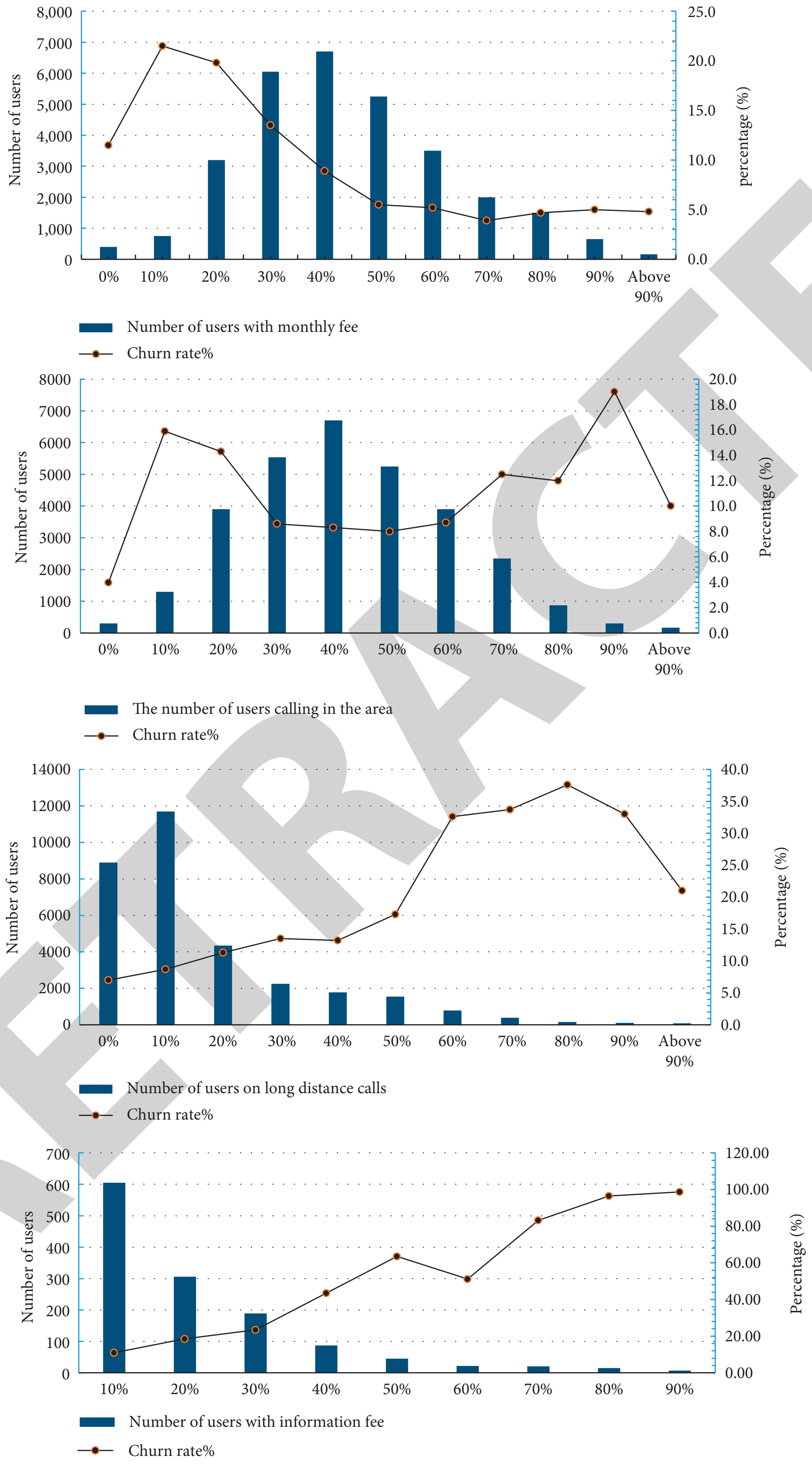

Figure 10: Relationship between customer churn and four cost factors.

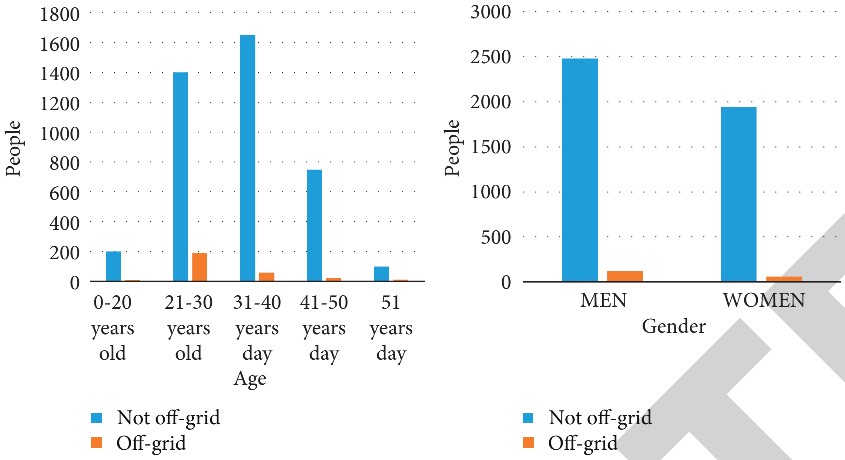

Figure 11: Statistical customer churn distribution based on age and gender.

proportion of monthly rental expenses increases. The distribution of the number of users in the area of call charge also basically obeys the normal distribution. However, the churn rate shows a trend of "low in the middle and high at both ends". Most users do not have long-distance call charges, or the long-distance call charges account for less than 10%. If it ignores the part that accounts for more than 80% (the sample size of this part is small, the error of the churn rate may be large), the churn rate of users generally increases with the increase of the proportion of long-distance fees. Although only a relatively small number of users incur information costs, the churn rate of these users is relatively high, higher than the average churn rate of 10.3%. And as the proportion of information fees increases, the churn rate also increases.

3.3. Establishment of Customer Churn Early Warning System Based on Data Mining.

The mining of customer data here is based on the consideration of the age of different customers and the loss of gender of different customers.

When the analysis results using age group as an indicator are shown in Figure 11, we can infer that there is a large difference in the off-grid rate of different age groups. It can be seen that age is an important factor in customer churn. Among them, it can be seen that the young and middle-aged groups have a high frequency of replacements and belong to users who are prone to off-grid. Younger people and middle-aged and elderly people are almost never lost and belong to long-term online users. The figure shows that customer churn is linked to gender. Compared with the results of previous analyses using age as an indicator, the associations between them are relatively small and their effects are small. It can be concluded that gender is not a key factor in customer churn factors, and its weight is small. Next is the customer churn situation in terms of monthly customer call time and caller ratio, respectively.

According to Figure 12, it can be seen that the off-grid rates of different time segments are significantly different. Therefore, it can be inferred that there is a certain correlation between the monthly call duration and the loss of customers. It is impossible to lose people who have an average call time of more than one hour per month. For users whose total call time per month is less than 1 hour, there is a high dropout rate. Such users may use multiple cards at the same time or for other reasons. Finally, the average monthly call charges of customers and the distribution of customers churn statistics from different traffic angles.

It can be seen from Figure 13 that the longer the network access time, the lower the corresponding customer churn rate level will be. Customers with high traffic demand are more likely to choose packages with cheaper tariffs and more traffic.

Table 4 shows the impact of network access duration index results on customer churn.

3.4. The Established Model

(1) The experiment uses the software SPSS 19.0 to perform binary logistic regression analysis on the customer data set. The final results of the experiment are shown in Table 3:

After processing the above data, the relational expression of churn = 0.254 ∗ length-1.354 can be obtained. Then its predictive performance needs to be examined. The accuracy rate here represents the success rate of relational prediction, which is an important basis for measuring the success of the model. The final function expression relation:

$$\begin{aligned}\text{churn} &= 0.041 * \text{mou} - 3.65 * \text{NetCall} - 1.25\\ &\quad * \text{MTraffick} + 0.28 * \text{length} - 1.75.\end{aligned} \tag{21}$$

The customer churn is 1.738, so the groundwork for the modeling is almost done. Its determinants and relative weights can be seen from the relational expressions.

(2) Model evaluation

Table 5 is the initial use of the preprocessed training set data to test the model. Therefore, in summary, we judge that the performance of the churn prediction model established

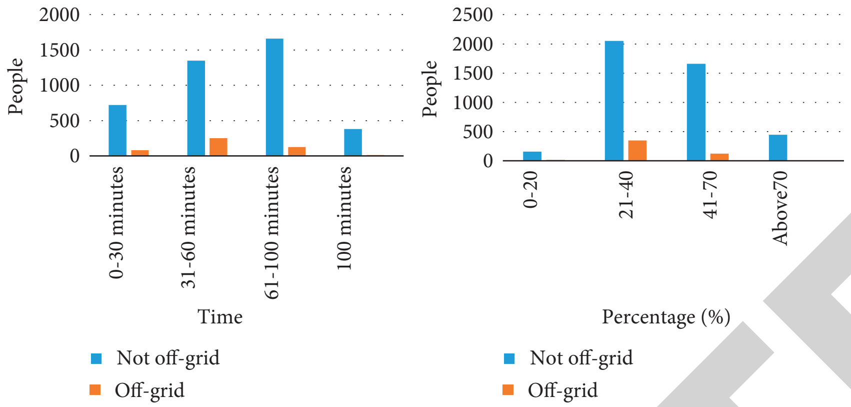

FIGURE 12: Statistics on the distribution of customer churn based on monthly customer call duration and caller ratio.

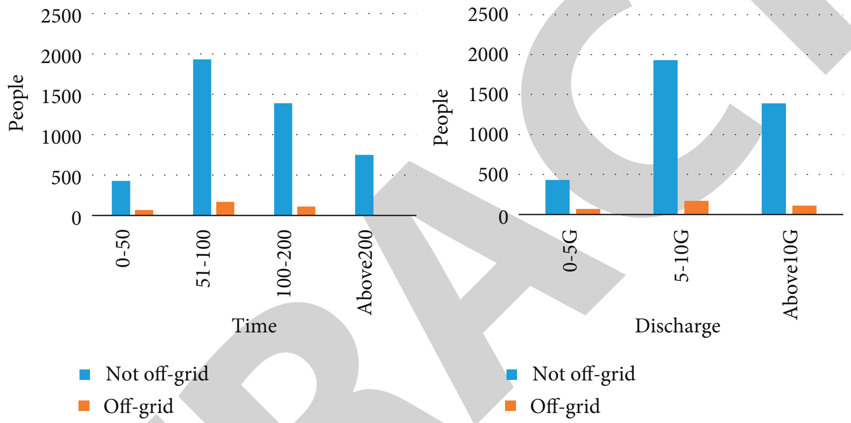

FIGURE 13: Statistics on customer churn based on average monthly call charges and different traffic.

TABLE 3: Logistic regression results.

Valuables not in the formula			
	Score	Df	Sig
Age	2.000	1	0.088
Sex	2.250	1	1.076
Mou	12.396	1	0.000
Initcall	421	1	0.051
MTraffick	33.7	1	3.000
Length	41.5	1	2.000

TABLE 4: Results of network access duration indicators.

	B	B.E	Wald	Of	Sig	Exp (B)
Length	−1.254	0.040	39.605	1	000	1 290
Constant	−1.523	0.431	12. 490	1	000	1.218

TABLE 5: Logistic regression prediction accuracy.

	Off-grid.	Not off-grid
Model predicts churn	251	4359
Actual off-grid population	432	4168
Prediction accuracy	58.1%	95.6%

in this paper is reliable. It thus finally establishes a customer churn prediction model. The average accuracy of its prediction is 75.4%, which is relatively successful [22].

4. Discussion

The application of the decision tree model is conducive to the mining of marketing value, and at the same time, it mines the data of potential purchasing customers. After training the decision tree, it classifies the customer information according to the customer's value, to improve the enthusiasm of employees and the success rate of marketing.

The vigorous development of mobile communication technology has made customer data resources one of the important resources today. In the face of such a large amount of data resources, enterprises should try their best to give full play to their own initiative.

This paper also uses logistic regression as a control experiment. This paper has adopted two different ways to model. Controlled experiments explore indicators that have an impact on customer churns, such as call duration, internetwork call ratio, traffic usage, average monthly consumption, and network access time. It then adopts the logistic regression method on the customer data and finally deduces a linear function relational expression. Experiments show that the functional relationship can be used to predict the accuracy of 79.5% of the test set. The RFT model proposed in this paper fully considers the key factor of customer value. Its accuracy can reach 87.5%, 86.6%, 82.5%, 86.4%, 83.7%, 82.2%, 87.8%, and 83.9%. Compared with the 79.2% obtained by using the logistic regression model directly in the control group, the prediction accuracy has been greatly improved.

After careful analysis, we found that the reason the data of the traditional operational system cannot support the analytical system, such as the customer churn warning system, is that the operational system does not save historical data. The current situation is that, with the development of enterprises, systems such as billing, business, and accounting are still operational, but they also save a certain amount of historical data. Its just that these data are not on a unified platform. In a sense, these systems have formed several data parts. Therefore, these data marks can be used as a basis. As long as the data is effectively integrated, the basic requirements of the application of the customer churn early warning system can be met.

5. Conclusions

With the increasing development of today's mobile communication technology, everyone generates network data every day. Therefore, the application of data mining plays an important role in dealing with customer churn. It is also a hot issue in the current research on forecasting models, improving the accuracy of forecasting models and enhancing interpretability. Some of its research has been applied to telecommunication services, financial insurance, passenger transport services, and other industries. It improves the management level of customer loss and carries out effective customer retention to achieve orderly market competition of enterprises and effective supervision by regulatory authorities. This is of great significance to the stability of the company's development. The predictability of customer churn is also critical to the importance of timely adjustments for businesses.

Data Availability

The data underlying the results presented in the study are available within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] R. J. Oskouei, N. M. Kor, and S. A. Maleki, "Data mining and medical world: breast cancers' diagnosis, treatment, prognosis and challenges," *American journal of cancer research*, vol. 7, no. 3, pp. 610–627, 2017.

[2] J. Zhang, "Interaction design research based on large data rule mining and blockchain communication technology," *Soft Computing*, vol. 24, no. 21, pp. 16593–16604, 2020.

[3] Q. Yang and S. Y. Kuang, "Abnormal data mining technology in remote virtual education communication," *Shenyang Gongye Daxue Xuebao/Journal of Shenyang University of Technology*, vol. 39, no. 4, pp. 412–416, 2017.

[4] L.-N. Wang, G.-M. Tan, and C.-R. Zang, "A network method to identify the dynamic changes of the data flow with spatio-temporal feature," *Applied Intelligence*, vol. 52, no. 5, pp. 5584–5593, 2021.

[5] Yumurtaci, "A Re-evaluation of mobile communication technology: a theoretical approach for technology evaluation in contemporary digital learning," *The Turkish Online Journal of Distance Education*, vol. 18, no. 1, p. 213, 2017.

[6] A. Kišić, "Information and communications technologies as a driver of effective internal communication," *Open Journal for Information Technology*, vol. 3, no. 2, pp. 39–52, 2020.

[7] E.-B. Lee, J. Kim, and S.-G. Lee, "Predicting customer churn in mobile industry using data mining technology," *Industrial Management & Data Systems*, vol. 117, no. 1, pp. 90–109, 2017.

[8] T. Y. Fei, L. H. Shuan, L. J. Yan, and S. W. King, "Prediction on customer churn in the telecommunications sector using discretization and Naïve Bayes classifier," *International Journal of Advances in Soft Computing and Its Applications*, vol. 9, no. 3, pp. 23–35, 2017.

[9] S. A Amatare and A. K. Ojo, "Predicting customer churn in telecommunication industry using convolutional neural network model," *IOSR Journal of Computer Engineering*, vol. 22, no. 3, pp. 54–59, 2021.

[10] A. Papa, Y. Shemet, and A. Yarovyi, "Analysis of fuzzy logic methods for forecasting customer churn," *Technology Audit and Production Reserves*, vol. 1, pp. 12–14, 2021.

[11] L. Almuqren, F. S. Alrayes, and A. I. Cristea, "An empirical study on customer churn behaviours prediction using Arabic twitter mining approach," *Future Internet*, vol. 13, no. 7, p. 175, 2021.

[12] M. Zhao, Q. Zeng, M. Chang, and Q. J. Tong, "A prediction model of customer churn considering customer value: an empirical research of telecom industry in China," *Discrete Dynamics in Nature and Society*, vol. 2021, no. 5, pp. 1–12, Article ID 7160527, 2021.

[13] I. V. Pustokhina, D. A. Pustokhin, P. T. Nguyen, and M. K. Elhoseny, "Multi-objective rain optimization algorithm with WELM model for customer churn prediction in telecommunication sector," *Complex & Intelligent Systems*, vol. 12, pp. 1–13, 2021.

[14] T. W Cenggoro, R. A. Wirastari, E. Rudianto, M. I. Mohadi, and B. Pardamean, "Deep learning as a vector embedding model for customer churn," *Procedia Computer Science*, vol. 179, no. 7, pp. 624–631, 2021.

[15] B Senthilnayaki, M Swetha, and D Nivedha, "Customer churn prediction," *Iarjset*, vol. 8, no. 6, pp. 527–531, 2021.

[16] J. U. Becker, M. Spann, and Barrot, "Impact of proactive postsales service and cross-selling activities on customer churn and service calls," *Journal of Service Research*, vol. 23, no. 1, pp. 53–69, 2020.

[17] S Venkatesh and M Jeyakarthic, "Metaheuristic based optimal feature subset selection with gradient boosting tree model for IoT assisted customer churn prediction," *Seybold Report*, vol. 15, no. 7, pp. 334–351, 2020.

[18] B. Yüceoğlu, "The effect of the length of the customer event history and the staying power of the predictive models in the customer churn prediction: case study of migros sanal market," *Academic Platform Journal of Engineering and Science*, vol. 8, no. 3, pp. 450–455, 2020.

[19] S. Venkatesh and D. M. Jeyakarthic, "Adagrad optimizer with elephant herding optimization based hyper parameter tuned bidirectional LSTM for customer churn prediction in IoT enabled cloud environment," *Webology*, vol. 17, no. 2, pp. 631–651, 2020.

[20] A. Deligiannis, C. Argyriou, and C. Argyriou, “Designing a real-time data-driven customer churn risk indicator for subscription commerce,” *International Journal of Information Engineering and Electronic Business*, vol. 12, no. 4, pp. 1–14, 2020.

[21] J. Sawant, “A study on customer churn in the telecommunications industry,” *International Journal of Management*, vol. 08, no. 03, pp. 121–124, 2020.

[22] K. Eria and B. P. Marikannan, “Significance-based feature extraction for customer churn prediction data in the telecom sector,” *Journal of Computational and Theoretical Nanoscience*, vol. 16, no. 8, pp. 3428–3431, 2019.